DOCUMENT RESUME

ED 299 924 HE 021 918

AUTHOR	Astin, Alexander W.; And Others
TITLE	Three Presentations: From the Third National Conference on Assessment in Higher Education. (Chicago, Illinois, June 8-11, 1988).
INSTITUTION	American Association for Higher Education, Washington, D.C.
PUB DATE	88
NOTE	57p.; Document collected as part of the American Association for Higher Education Assessment Forum.
AVAILABLE FROM	AAHE Assessment Forum, American Association for Higher Education, One Dupont Circle, Suite 600, Washington, DC 20036 ($10.00).
PUB TYPE	Viewpoints (120) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Accountability; *Educational Assessment; Higher Education; Incentives; Instructional Effectiveness; Multiple Choice Tests; *Outcomes of Education; Productivity; Role of Education; Standardized Tests; *Student Development; Testing
IDENTIFIERS	*AAHE Assessment Forum; *College Outcomes Assessment; Florida

ABSTRACT

Three presentations from the Third National Conference on Assessment in Higher Education are included. In "Assessment and Human Values: Confessions of a Reformed Number Cruncher," Alexander W. Astin, focuses on measuring education productivity, assessment lessons from the Cooperative Institutional Research Program, a talent-development model of excellence, assessment and values, multiple-choice tests, holistic methods, assessing affective outcomes, and beyond narcissism. It stresses that the key to achieving institutional transcendence is ultimately in how excellence is defined. "Assessment and Incentives: The Medium is the Message" (Linda Darling-Hammond) discusses from a teacher's point of view the following: how measurement changes behavior; incentives: a parable with lessons; the K-12 experience;, limits of standardized testing; effects of testing on teaching and learning; and policy making and assessment. Important factors are educating those who would impose hasty or inadequate methods, and insisting on intellectual honesty and educational validity. "The Assessment Movement: What Next? Who Cares?" (Robert H. McCabe) gives a community college president's views on access and standards, the public call for accountability, state initiatives in assessment, the Florida experience; and institutional assessment initiatives. The assessment movement is growing in tandem with the teaching/learning movement and can be considered an element of it. The future of assessment is in improving student development through more effective teaching and learning. (SM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED299924

Three Presentations:

FROM THE THIRD NATIONAL CONFERENCE ON ASSESSMENT IN HIGHER EDUCATION

JUNE 8–11, 1988 CHICAGO

Alexander W. Astin
Linda Darling-Hammond
Robert H. McCabe

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

AAHE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 021 918

THE AAHE ASSESSMENT FORUM

AMERICAN ASSOCIATION FOR HIGHER EDUCATION

One Dupont Circle
Suite 600
Washington, D.C. 20036
202/293-6440

AMERICAN ASSOCIATION
FOR HIGHER EDUCATION

Board of Directors

Chair
Joseph F. Kauffman
University of Wisconsin
Madison

Chair-Elect
Adele S. Simmons
Hampshire College

Vice Chair
Reatha Clark King
Metropolitan State University

Past Chair
Harriet W. Sheridan
Brown University

Carlos H. Arce
NuStats, Inc.

Estela M. Bensimon
Teachers College
Columbia University

Anne L. Bryant
American Association of
University Women

Donald L. Fruehling
McGraw-Hill, Inc.

Ellen V. Futter
Barnard College

Jerry G. Gaff
Hamline University

Zelda F. Gamson
University of Michigan

Stephen R. Graubard
Daedalus

Joseph Katz
State University of New York
at Stony Brook

Arthur E. Levine
Bradford College

Frank Newman
Education Commission
of the States

Alan Pifer
Carnegie Corporation
of New York

W. Ann Reynolds
The California State
University

Piedad F. Robertson
Miami-Dade Community
College

D. Wayne Silby
Groupware Systems

P. Michael Timpane
Teachers College
Columbia University

President
Russell Edgerton

The AAHE ASSESSMENT FORUM is a three-year project supported by the Fund for the Improvement of Postsecondary Education. It entails three distinct but overlapping activities:

--an annual conference
(the first scheduled for June 14-17, 1987, in Denver)

--commissioned papers
(focused on implementation and other timely assessment concerns; available through the Forum for a small fee)

--information services
(including consultation, referrals, a national directory, and more)

This paper is part of an on-going assessment collection maintained by the Forum. We are pleased to make it more widely available through the ERIC system.

For further information about ASSESSMENT FORUM activities, contact Patricia Hutchings, Director, AAHE ASSESSMENT FORUM, One Dupont Circle, Suite 600, Washington, DC 20036

BEST COPY AVAILABLE

ASSESSMENT AND HUMAN VALUES: CONFESSIONS OF A REFORMED NUMBER CRUNCHER

An address by:

Alexander W. Astin
UCLA Graduate School
of Education

Presented at:

The AAHE Assessment Forum
Third National Conference on Assessment
in Higher Education
June 8-11, 1988
Chicago, Illinois

Assessment and Human Values: Confessions of a Reformed Number-Cruncher*

Alexander W. Astin
University of California
Los Angeles

My talk this morning is going to be a mixture of technical stuff on assessment and a little philosophy. Since what I have to say will probably make more sense if it also includes some autobiographical notes, so let's start first with my graduate education. My doctorate in psychology was based on a double major--counseling psychology and quantitative psychology--and to an extent this dual emphasis represents two sides of myself that for the past thirty years have been struggling to get in balance and in tune with each other. My interest in mental health and counseling and psychotherapy is what initially got me into psychology, but in graduate school I quickly found that I liked writing and that statistics and research design came very easily to me. Research was just plain fun. Counseling and psychotherapy, on the other hand, not only seemed to be much more difficult, but I also came to have some serious doubts about whether I was really doing my clients and patients very much good.

Although my first postdoctoral job was as a clinical psychologist, I managed to find time to do some research and writing. And although my supervisor seemed to feel that I was a good clinician, after two years I decided to look for a full-time research position.

How I eventually ended up in the field of <u>educational</u> research is a complicated story that's probably not very relevant to our topic today; suffice it to say that I quickly found higher education to be a fascinating and challenging field where the problems seemed--on the surface, at least--to be much more <u>tractable</u> than those in the mental health field.

Measuring Education Productivity

My first higher education research was concerned with something called "Ph.D. productivity." Researchers at Wesleyan University and the University of Chicago (Knapp and

*Presented at the AAHE Assessment Conference, Chicago, June 8-11, 1988.

Goodrich, 1952; Knapp and Greenbaum, 1953) had found that certain colleges were much more likely than others to produce graduates who eventually went on to win graduate fellowships and to earn the Ph.D. degree. Since the "highly productive" colleges also tended to have larger libraries, better student-faculty ratios, and more faculty who themselves had Ph.D.'s, the researchers concluded that these superior facilities and resources were somehow responsible for the colleges' higher productivity.

I was working at the National Merit Scholarship Corporation at the time, and noticed that the "highly productive" colleges tended to be the same ones that the Merit Scholars preferred to attend. This fact prompted me to ask a rather simple question: Could a college's "output" of Ph.D.s be explained simply in terms of its initial "input" of talented freshmen? To test this possibility we conducted a series of studied which showed that, as far as "Ph.D. output" is concerned, the "student input" is by far the most important factor. Indeed, it turned out that, when you took student inputs into account, some of the so-called "highly productive" institutions were actually underproducing Ph.D.s, whereas some of those with more modest outputs were actually producing more than one would expect from their student input (Astin, 1961, 1963).

These early studies were critical in teaching us three fundamental lessons about assessment in higher education. Let me briefly summarize these lessons for you (Fig 1.).

1. First, the "output" of an institution-- whether we measure this in terms of how many graduates earn advanced degrees, how much money the alumni earn, or whatever--doesn't really tell us much about its educational impact or educational effectiveness. Rather, outputs must always be evaluated in terms of inputs. (Fig. 2) This is a particularly important principle for American higher education, given the fact that the three thousand institutions in our system differ so greatly in the kinds of students they enroll. I am

Note: Figures refer to slides; not included here.

speaking here, of course, of the need for longitudinal data. You can also see the relevance here to the value-added concept.

2. Second, an output measure such as Ph.D. productivity is not determined solely by a single input measure such as student ability. On the contrary, even in our earliest studies of this phenomonon we found that input variables such as the student's sex and major field of study are at least as important as ability in determining Ph.D. outputs.

3. Third, even if we have good longitudinal input and student output data, our understanding of the educational process will still be limited if we lack information on the college environment. (Fig. 3) Thus, it is one thing to know that your college overproduces or underproduces Ph.D.'s, but quite another to understand why. What is it, in other words, about the environment of a college that causes it to over- or under-produce? This last lesson suggests that input and output data, by themselves, are of limited usefulness. Rather, what we need in addition is information about the students' educational environment and experience: the courses, programs, facilities, faculty, and peer groups to which to which the student is exposed.

Perhaps the need for these three kinds of data can be better understood with an analogy from the physical science of astronomy. Having only output data would be like taking a single shapshot of a heavenly body, say the sun or the moon or some planet. We might be able to determine its distance and size, but not much else. Having a series of snapshots of the same heavenly body would represent an improvement, because we could chart how it changes over time.

This is analogous to having both input and output data. But limiting ourselves only to input and output data would be like having a good description of the movement of the heavenly body without knowing what was happening to any of the other planets or satellites or stars in its immediate environment. In other words, simply being able to document changes in the behavior of a group of students over a period of time is of limited value if you don't know what forces were acting on these students during the same period of time.

Perhaps an even better analogy can be found in the field of health care. Let's say we were trying to enhance our understanding of how best to treat patients in a hospital. Imagine how difficult it would be if all we did was to collect output information on how long each patient stayed, and whether they got better, got worse, or died. How could we expect to learn much about how best to care for our patients if we don't know which patients got which therapies, which operations, or which medications? This is the equivalent of studying student development with no "environmental" data on what courses they took, where they lived, how much they studied, and so on. But if all we do is to collect environmental and outcome (post-test) data, we won't even know which students--in terms of high school background, family background, major, interests, and so--gained the most and which gained the least. This is the equivalent of studying patients' treatments and eventual outcomes without knowing anything about their history or diagnosis at the point of entry.

These early studies present a number of challenges to a quantitatively-oriented researcher. How, for example, do you control for student inputs and how do you assess environmental effect? These are extremely complex and difficult technical issues that we don't have time to cover today. Another quantitative challenge is how best to measure student outputs: What are the important and relevant outputs and how do we assess them? (I will return to this question shortly.) But perhaps the most difficult assessment question of all is how to measure the educational environment. This is not only a complex and important problem, but it is clearly our most seriously neglected problem. Many of us in the assessment field have agonized a good deal about what student

outcomes we want to assess or the merits of pretesting and post-testing, but few of us have given much thought to the problem of assessing and documenting the students' educational experiences.

At the National Merit Scholarship Corporation we were fortunate in having access to a large group of highly talented and cooperative students who were willing to provide us with input and output data via mailed questionnaires and personality inventories. We had access to as many as 35,000 highly able new students every year. It was during this time that I began to do some serious number crunching--first with IBM punched card equipment and subsequently with what later came to be called "first generation" computers. The old vacuum tube jobs. My colleague Bob Nichols--God bless him--virtually coerced me into letting him teach me the FORTRAN language--and before long we were writing some pretty fancy miltivariate programs for analyzing large data bases. Some of our data runs were bigger than they had ever had at the Northwestern University computer center, and on a few occasions we crunched so many numbers so fast that we would actually fatigue one of the vacuum tubes to the point where it would quit working temporarily.

But the National Merit Finalists represented such a small fraction of the student population--and such a highly biased fraction at that--that we soon decided to go for broke and to survey <u>all</u> of the freshmen at a random sample of 248 four-year colleges and universities. We crammed about 20 items of input information onto 5 x 8 cards which were filled out by 127,000 freshmen in the fall of 1961. This was designed basically as a survey of student input characteristics which we eventually hoped to follow up longitudinally. Much of the data, such as the students' choices of major fields and careers, had to be coded by hand, and <u>all</u> of it had to be key-punched. We had so much data that in trying to read the input file we sometimes exceeded the reliability of the card reader or the tape drive. In those days the computer systems' software was such that if you had a read error from the input file, the whole job was aborted--so these errors were very costly.

I have digressed a bit to talk about our massive data processing problems to make a point: From our early experience with the Ph.D. productivity problem we came to realize several important truths about multi-campus studies of college student development. First, you need a large number of institutions in order to represent adequately the great diversity of college

environments in the United States. Second, you need to measure a large number of variables, not only to reflect the many different types of student outcomes, but also to make sure that you have controlled for most of the potentially biasing student input variables. Finally, you need large numbers of students in order to perform sophisticated analyses of the many input, output, and environmental variables. None of this would have been possible, of course, without digital computers and, eventually, without optical scanners.

Assessment Lessons from CIRP

In those days we felt we were really engaging in research that was on the cutting edge of higher education, and the scientific and technical journals were very receptive to our studies. What was somewhat frustrating, however, was that the higher education community didn't seem to be paying much attention to our findings, even though it seemed to us that what we were coming up with had a lot of relevance to what college faculty and adminisrators were supposed to be doing. It was primarily for this reason that in 1965 I jumped at the chance to move to the American Council on Education (ACE) in Washington to establish a research office. Since the Council is where the top administrators and policy makers in higher education congregate, I felt, on the one hand, that these leaders would benefit from a better knowledge of how their institutions were actually affecting their students, and that we, on the other hand, could benefit from their advice and counsel in conducting this research.

These explanations eventually proved to be somewhat naive, as I will shortly point out. In any case, one of our first major activities at the Council was to set up the Cooperative Institutional Research Program (CIRP). As you might guess from our previous experience at the National Merit Scholarship Corporation, CIRP was conceived of as a longitudinal study of how students are affected by their college environments. We started out in the fall of 1966 surveying the entering freshmen at a representative sample of 300 two- and four-year institutions and published our first national norms on American college freshmen. (Fig. 4) An interesting sidelight to this is that we initially produced these norms merely as an incentive for institutions to participate. Our main goal, of course, was to conduct longitudinal studies of students by following up the entering freshmen,

but we were concerned that institutions would not be very interested in participating if they had to wait several years for feedback from the longitudinal follow-ups. The freshman norms provided each institution with something they could use right away. Of course, after 22 years these freshman surveys have acquired an identity and a life of their own; as a matter of fact, the freshman survey is probably the most visible thing we do now and many people are not even aware of our follow-ups and of the basic longitudinal design of CIRP.

As we continued with our freshman surveys and the longitudinal follow-ups, it gradually began to dawn on us that virtually everything we were learning about assessment and research from our multi-institutional studies could be applied with equal validity to assessment activities at an individual institution. Let me briefly review some of these principles (Fig. 5):

1. The need for multiple outcome measures. Clearly, no single institution's programs and impact on students can be adequately assessed with a single outcome measure.

2. The need for input assessments. It is not very useful simply to know how students perform at the exit point. In addition to providing a basis for measuring growth and change, student input data can enhance our understanding of the entering students' background, talents, aspirations, and educational needs. Such information can be extremely useful in program planning.

3. The need for environmental data. Just as students who attend different institutions are often confronted with quite different types of educational programs, so are the students at a single institution often exposed to quite different kinds of educational experiences. Among the more obvious environmental differences are the majors students pick, the particular courses they take, and the particular professors or advisors who counsel them. But there are many other variations and environmental experiences that can make a substantial difference in how students actually develop: (Fig. 6) where they live while attending college, how they study, how they use their time, how they support themselves, what kinds of organizations they join, what kinds of co-curricular activities they participate in, how much and what kinds of contact they have with faculty outside of class, and whether or not they participate in special educational activities such as honors

programs, developmental education, cooperative education, study abroad, and so on. These are the things we can directly control on a campus, and any or all of them can make a difference in how students change or develop during the college years. Perhaps our main job as assessment specialists on the campus is to help our students and our colleagues gain a better understanding of how these various environmental factors affect different student outcomes. But note that, in order to do this, we need to creata a data file which incorporates all three kinds of information in a single place. (Fig. 7) Without such a data base there is simply no way that our outcomes assessments can be used to determine (a) why some students develop so differently than others; (b) what types of programs and experiences work best for what types of students; and (c) which aspects of our institutional environment should be preserved and strengthened and which should be changed. I might also add here that the findings from our FIPSE-funded value-added consortium (which is making a presentation on Saturday morning) show clearly that the biggest single obstacle to implementing an effective program of outcomes assessment is the absence of a comprehensive student data base which incorporates input, output, and environmental data on individual students in the same place.

During the early years of the CIRP we conducted several longitudinal follow-ups and publishd a number of books and articles on how students are affected by the type of college they attend. And even though this work was very satisfying and rewarding intellectually and professionally, something was missing. In retrospect it seems that the clinical psychologist or do-gooder part of my personality had been suffering from neglect. In order to give greater expression to this side of myself, I and my colleagues began to use CIRP data to explore a number of contemporary social problems and issues. Over the years we produced books on the disadvantaged student (H. Astin et al, 1970), campus unrest (Astin et al, 1975) open admissions (Rossmann et al, 1974), college dropouts (Astin, 1975), and ethnic minorities (Astin, 1982). We even did one for the Carnegie Commission on what we called The Invisible Colleges (Astin and Lee, 1971): the poorest, smallest colleges that no one seemed to care much about.

A Talent-Development Model of Excellence

In spite of all this activity, it gradually began to dawn on me that many of the educational leaders who participated in the Council's activities had very little interest in our studies of student development and institutional impact. Rather, what seemed to preoccupy the attentions of college and university heads was how to bring more resources into their institutions and how to preserve and enhance their institution's reputations. These impressions were what led me eventually to conclude that our traditional thinking about "excellence" in higher education is dominated by two views: (Fig. 8): excellence as resources, and excellence as reputation (Astin, 1975). As some of you know, for several years now I have been very critical of these traditional views of excellence and have suggested an alternative approach which I call the "talent development" view. Under a talent development approach, excellence is not necessarily dependent on what resources or reputation you have, but rather on how effectively you educate your students. I realize that some of our colleagues have been critical of the talent development or value-added approach, and for the life of me I can't see what all the fuss is about. Without getting into any of the details of this debate, let me just say this: To me, what education is really all about is learning, change, growth, and development. This is all that is implied by the talent development concept. How you decide to measure change is up to you. But if you don't think that education is fundamentally concerned with change and growth and development in students, then the talent development concept may indeed not suit your purposes.

The debate over how we ought to define "excellence" is really a debate over values. Given that the resources and reputational approach lead us to compete for higher and higher positions in the institutional pecking order, the values underlying these traditional views are inherently competitive, materialistic, and perhaps even narcissistic. The talent development approach, on the other hand, requires us to do everything we can to contribute to the intellectual and personal development of our students. Clearly, the implicit values underlying this approach are more cooperative and altruistic than competitive and materialistic.

What many of us fail to realize is that these same values have tremendous implications for how we approach assessment in higher education. I have lately come to believe that the values of an institution are reflected in the kinds of information it collects and pays attention to. Under the reputational and resource approaches, for example, we are primarily interested in input data on the entering student, since their high school GPAs and admissions test scores are viewed as a measure of our excellence. We are excellent merely by having bright students, not necessarily by educating them well. Under these traditional views we might also be interested in output data on the alumni: how much they earn, how many are listed in Who's Who, etcetera, as another index of our excellence. If our graduates are excellent, then certainly we must be excellent. Under a talent development view, however, neither input nor output data, by themselves, is very useful. Rather, we really need both kinds of data, because what we seek to find out is how much students are actually learning and how they are developing over time.

Assessment and Values

But lately I've come to realize that these value questions go far beyond merely collecting input and output data and trying to understand how much and how well our students are learning. An equally critical value question concerns what aspects of student development we decide to assess, and how we choose to assess them. Whether we like it or not, there is no such thing as value-free assessment. Our very choice of measures--our decision to measure some things and not others--involves value judgments.

Take, for example, the familiar dichotomy between cognitive and affective outcomes. Most of us are inclined to shy away from assessing affective outcomes because we think they are too value-laden. We feel much more comfortable limiting our assessments to cognitive outcomes. College, after all, is supposed to develop the student's intellect, so how can we go wrong if we focus on cognitive variables? But wait a minute. If you read through a few college catalogs, you begin to realize that colleges claim to be concerned about such "affective" things as good judgment, citizenship, social responsibility, character, and the like. Indeed, most descriptions of the "liberally educated" person that I've ever read sound at least as affective as they do cognitive. I

want to return to the issue of affective outcomes shortly, but first let's look a little more closely at the cognitive side.

By far the toughest assessment challenge in the cognitive domain is to measure talent development in the area of general education. Basically what we try to do here is to assess those qualities that the G.E. program is supposed to foster and develop.

For some institutions this might include basic comunication abilities such as skill in written composition. I have also seen institutions try to assess reading and even speaking skills, but I've almost never heard of a program that includes skill in listening. (The pioneering program at St. Edwards University in Austn, Texas is a notable exception.) What are the implicit values involved in such choices? Why do we emphasize good writing and speaking so much more than good listening? What values are involved in such choices?

Our basic purpose in assessing G.E. outcomes is, of course, to reflect the intent of the curriculum. Yet the G.E. curriculum--whether it consists of core courses or distributional requirements--is itself a statement of values. If you doubt this, I simply refer you to the recent flap over Stanford University's revised G.E. curriculum. After much internal controversy and debate, the Stanford faculty decided to change the required reading list to include more material by and about women, minorities, and non-western societies. Education Secretary Bennett injected himself into this controversy by being strongly critical of Stanford's decision, accusing the Stanford faculty of selling out the great books tradition to transient political pressures. Regardless of whose views you endorse on this matter, how can you possibly read about this controversy and continue to believe that any curriculum is not heavily value-laden?

Most of us seem comfortable in trying to asess the cognitive skills that our core curriculum is designed to foster--whether these be computational skills, verbal communication skills, computer skills, or critical thinking skills. The reason why we usually do not question such decisions, I think, is that we tend to believe that focusing on what are in effect vocational skills is not a value issue. After all, who would argue about the usefulness of equiping students with skills that will enable them to perform better in their jobs? Are not these skills the same ones that are

most valued by the companies that employ most of our graduates? But what about the other aspects of the students' life: their family, their friends, their role as citizens, and their participation in the comunity? Is it not a statement of values to decide to focus on talents that are of interest to employers and to neglect talents that are relevant to the many other aspects of the students' life after college?

The Multiple-Choice Tests

Let's now consider how we assess cognitive outcomes. How value-free are our assessment methods? Our old friend, the multiple-choice test, is still by far the most popular approach. Multiple-choice tests are popular for at least two reasons: They can be administered and scored very cheaply in large groups, and they naturally yield quantitative scores that make it easy to differentiate among students. Aside from the many technical criticisms that one can make against this multiple-choice methodology, I have several concerns that are more value-oriented.

First is the way multiple-choice tests are scored. Typically, the number of right answers (or a weighted combination of rights minus wrongs) is converted into some type of normed score, either a percentile or a standard score. Now what do we really do when we make such a conversion? We have lost the basic information about how many items (and which ones) the student got right and wrong, and replaced this information with a score indicating only how well the student performs in relation to other students. By using tests that are scored normatively, we are basically putting students in competition with each other. So the implied value here would seem to be that the cognitive performance of any given student should be judged only competitively: How much better or worse did the student do in when compared to other students? This competitive scoring procedure is essentially identical in spirit to traditional classroom grading, especially if the grading is done "on the curve." I might add that these relativistic and competitively-scored tests are difficult to use in assessing talent development, because they make it virtually impossible to determine how much a student has actually changed or improved over time. All we can say is that the student's performance has increased or decreased in relation to other students.

There is another, perhaps even more subtle problem with normative assessment, whether it be through letter grades or standardized tests: When we choose to assess performance using a normed instrument, we create what the economists would call a "scarce good." Only so many students can be at the top of their class and only so many students can score above the 90th percentile. No matter how hard students work and no matter how much they actually learn, there will always be only so many "excellent" test scores and so many "excellent" grades. Normative assessment, in other words, automatically constrains how much "excellence" you can have. The important thing to realize is that this shortage is a completely artificial one, rather than something inherent in the trait being assessed. The shortage, in other words, is something created by the assessment method itself.

As with any scarce good, the scarcity itself tends to exaggerate the importance of being at the top, so that below average or even average performance is often viewed as failure. Normative scoring, in other words, guarantees that a substantial number of students, if not the majority, will view themselves as failures.

Before discussing the other limitations of multiple-choice tests, I would like to point out that there is something that we in the assessment field can do to overcome the negative consequences of most norm-referenced tests. Very simply, we can insist that the testing companies give us back the raw score results and, ideally, the results from individual test questions. Raw scores provide a way to measure how much each individual student is actually learning or improving over time, without requiring any competitive comparisons with other students. Furthermore, results from individual test questions can be useful to individual students in understanding their particular strengths and weaknesses. Item results aggregated across students can be very useful in curriculum planning and course evaluation. I feel srongly that all of us who utilize any type of standardized test in our assessment work should begin insisting that the testmakers give us this kind of feedback.

My second concern about multiple-choice tests is the artificiality of the task itself. After students finish their formal education, the ability to find a correct answer from a predetermined set

of alternatives has a very limited usefulness. How often in real life is any of us presented with a prepackaged set of possible answers to a question, only one of which is correct? And how often are we required to read the question and find the answer under intense time pressure? How often do life's problems take such a bizarre form? And what about the myriad real-life problems that call for creative solutions? My point here is that the ability to perform well on such tasks is so highly specialized and so foreign to the kinds of real-life problems that we normally confront that I really wonder if we educators have been wise to make such liberal and uncritical use of the multiple-choice test. The testmakers might respond that such tests have "predictive validity," and indeed they do. But in such validity studies the outcome being predicted is almost always school or college grades or simply another test constructed in the same manner!

This problem is perhaps best illustrated by the fact that the multiple-choice test is not well-suited to many important kinds of educational outcomes, but especially not to those that require the exercise of creative talents. Creativity can be expressed basically in two ways: through creative products, and through some sort of creative performance. (Fig. 9) Creative products include such things as essays, research papers, scripts, films, videos, works of art, and musical compositions. Creative performances include equally diverse activities such as public speaking, dance, musical recitals, and theater productions. Depending upon how broadly you choose to define "creativity," you might also include performance outcomes such as leadership behavior, public service, and athletics. Clearly, the multiple-choice test is an inappropriate technology for assessing many types of creative outcomes that are highly valued not only in the academic community but also in later life. Thus, when we insist upon putting our principal assessment emphasis on student outcomes that can be measured through multiple-choice tests, we are implicitly assigning low priority to student creativity.

My third value concern about normatively-scored multiple-choice tests is much more subtle and philosophical. It has to do wih the fact that when we administer and score such a test, we are implicitly placing the student along some kind of narrow continuum together with other students. In a sense, we are forcing all the square pegs and rectangular pegs and other oddly-

shaped pegs into the same round hole. This process does considerable violence to the rich and marvelous diversity that characterizes any group of students. To think that we can adequately capture this diversity through one or even several multiple-choice test scores is absurd.

My final concern about multiple-choice tests is the distance that they put between the student and the professor. Not only is the adminisration and scoring done impersonally, but the numerical feedback is dry and impersonal as well. Clearly, such a procedure is inimical to the close student-faculty contact that much of the research in higher education shows to be so important.

Holistic Methods

By this time you may wonder, "If Astin is so dubious about course grades and standardized multiple-choice tests of cognitive functioning, what would he put in their place?" Clearly, the student needs feedback of some kind, and the institution needs some way to document the student's progress. In my brief discussion of creative performance, I have already suggested that the individual creative products and performances of students probably have to be assessed in some kind of holistic fashion. The more I think about this problem, the more convinced I become that holistic feedback, whether written or spoken, is far and away the most powerful assessment tool we have for enhancing the educational process. Specifically, I am thinking here of the kinds of written feedback that students receive at places like Hampshire College, Alverno College, Empire State College,. and the University of California at Santa Cruz. If you ever had occasion to see such narrative feedback, especially from the individual student's perspective, you would appreciate its potential educational value. Not only is the feedback itself extremely informative and useful to the student, but since the process itself requires the professor to get to know the individual student's work personally, narrative feedback strengthens and enhances the relationship between student and faculty member.

This discussion not intended to imply that normed multiple-choice tests have no place in the assessment of cognitive development; rather, I simply wanted to challenge the widely-held belief that such tests are always the best method for assessing cognitive outcomes. Like every

other method, they have their advantages and limitations, and if one of the purposes of assessment is to enhance the teaching-learning process and to bring about greater contact between faculty and students, then narrative feedback has much to offer.

One problem with holistic written evaluations of student performance, of course, is that they do not necessarily yield numerical estimates of student performance to put in our data base and to use in our statistical analyses. This is a real problem for us number-crunchers. I would like to stress, however, that there is nothing inherent in the narrative or in any other qualitative assessment method process that precludes quantification, although not all institutions that use narrative feedback try to quantify their evaluations. One very simple approach would be to have the evaluator also complete a brief set of rating scales, with each scale representing a different skill or area of knowledge. Such an approach is not unlike the quantification involved in scoring essays or in judging musical and artistic compositions.

Another, perhaps more serious objection to holistic narrative evaluations is that they are highly labor intensive. If professors are going to be asked to undertake such evaluations of their students, what kinds of trade-offs are we going to make in terms of the professors' other job responsibilities? I believe that the best way to approach this question is first to recognize that we are once again dealing with a question of values. If we believe that students can benefit significantly from the experience of having a professor get to know their work well enough to write a detailed narrative evaluation of it, then what other, presumably less useful, activities can be traded off against the time required for the professor to carry out the evaluation and discuss it with the student? Each institution, of course, will have to answer this for itself, but it seems to me that one reasonable trade-off would be for professors to do a bit less lecturing or even to teach one less class. Not only would the students benefit from the personalized feedback, but the professors would probably welcome the variety introduced into their pedagogical activities. Certainly it behooves us to begin to study the potential efficacy of such tradeoffs.

Assessing Affective Outcomes

Let's return now to the issue of affective outcomes. I've already suggested that the importance of such outcomes is implied in the mission statements appearing in many college catalogs as well as in the notion of the liberally-educated person. What are some of these outcomes and how can they be assessed? In the area of affective skills we have a variety of potentially important qualities such as interpersonal competence, leadership ability, and empathy. The ability to empathize with others is, incidentally, probably dependent to some extent on one of the communications skills listed earlier: listening ability.

There are a number of other affective outcomes that seem to be relevant to the goals of a liberal education, although we do not ordinarily consider them to be "skills." These would include self-knowledge and self-understanding, maturity, motivation for further learning, understanding of other peoples and societies, self-esteem, social responsibility, and even good mental and physical health.

One affective area that needs more attention in our assessment activities is the student's own values. As some of you know, I have been involved in monitoring the values of incoming freshmen through the CIRP surveys for some 22 years now, and what I see happening troubles me. During the past two decades, students have become markedly more materialistic and more concerned with having power and status. (Fig. 10) They are increasingly coming to see an undergraduate education primarily as a means to make more money and less as a place to get a general education. (Fig. 11) At the same time, students have become less concerned about the well-being of others, the environment, and the community. (Fig. 12) These value changes have been accompanied by similar changes in the students' career plans, with careers in business reaching all-time highs in popularity, and careers in the human service occupations reaching all-time lows.

Lately the higher education community has begun to counteract some of these trends by creating a number of programs that are designed to encourage student participation in public and comunity service activities. The Campus Compact project, for example, is a consortium of some 130 institutions that are working together to establish community service programs for

undergraduates under the sponsorship of ECS. In my own state of California the legislature has passed a law requiring each campus of the university and the state university to establish some kind of volunteer for public service program for undergraduates.

In last year's report to the Board of Overseers, Harvard president Derek Bok said that "universities should be among the first to reaffirm the importance of basic values such as honesty, promise keeping, free expression, and nonviolence...[and] there is nothing odd or inappropriate...to make these values the foundation for a serious program to help students develop a strong set of moral standards." Bok also notes that "students must get help from their universities in developing moral standards or they are unlikely to get much assistance at all."

These trends would suggest that our political and higher education leaders are implicitly suggesting that social responsibility and concern for others is one of the qualities that higher education institutions should try to foster in their students. Certainly it would seem appropriate for those of us in the business of trying to assess student outcomes to join forces. Let's begin to introduce longitudinal measures of qualities such as empathy, concern for others, tolerance, social responsibility, and the like into our assessment programs.

Let me get just a bit more personal about this. As I look around me everywhere I see the great achievements of the intellect: atomic energy, genetic engineering, modern agriculture, modern medicine, and computers and other electronic marvels of every conceivable type. It is truly astounding. And at the same time I see the great affective and emotional and spiritual divisions that threaten our very existence: religious fanaticism and hatred, racial prejudice, nationalism and political divisions, widespread criminal behavior in the land of opportunity, and massive poverty and starvation in the face of unprecedented affluence. What this tells me is that is it time to redress the balance. It is time to begin shifting some of our educational interest and energy in the direction of our affective side -- to begin concerning ourselves much more directly with the development of beliefs and values that are going to heal our divisions, and which will help to create a society that is less materialistic and competitive and selfish and more generous and cooperative.

Beyond Narcissism

Let me close by returning once again to our different conception of excellence. In the long run, the kinds of assessments that we perform and how we use the results will depend heavily on our values, which means in part what conception of excellence we ultimately choose. During my days as a clinical psychologist I would have looked at the reputational and resource views as basically egocentric, in the sense that they emphasize what possessions we have and what others think of us. (Perhaps "narcissistic" would be even more accurate than "egocentric.") Under a talent development view, we identify ourselves instead in terms of what we do, what we contribute to others and what they in turn contribute to their communities and to the society. In other words, by adopting a talent development perspective, we, in effect, transcend out institutional egos to some extent and begin to view our institutions more in terms of their impact on the larger society.

The basic point to be made here is a rather simple one: When an institution exists primarily for its own sake, and when it identifies itself primarily in terms of its resources and reputation, its relationship with the society it is supposed to serve becomes exploitative and defensive, and its capacity to serve as an instrument for improving the society is compromised. In short, the biggest obstacle to higher education's serving as a major instrument for societal improvement is the institutional ego. In a sense, our colleges and universities need to learn how to transcend their institutional egos and to become more actively involved in what is going on in this society.

The notion of transcendence is, of course, a frequent theme in psychological theories. When I was doing psychotherapy, I was continually made aware of the fact that my ability to help a troubled patient was limited to the extent that I allowed my ego to become a prominent part of the therapeutic process. Egotistical therapists are less helpful to patients because their sense of worth comes primarily from the power they wield over the patients and from the patient's sense of helplessness and dependency. The most effective therapists, on the other hand, are those who are able to transcend their egos to the extent necsessary to emphathize fully with the patient and to create an accepting and supportive therapeutic clinical environment. The therapeutic focus, in other words, should be on the patient rather than the therapist.

I believe that there is a clear analogy here between patients and therapists, on the one hand, and students and institutions, on the other: The capacity for higher education to be a positive change agent in American society will depend upon our ability to transcend our institutional egos, our narcissism, and our self-interest, and to concern ourselves more directly with the impact we are having on our students and comunities.

I believe that the key to achieving this kind of institutional transcendence is in how we ultimately define our own excellence. Rather than continuing to see our excellence as limited to what we have (resources) or to our status and prestige (reputation), we need to see it more in terms of what we do and what we accomplish. And to do this we need to rely heavily on assessment. We need to know many things: How much and how well do our students learn? How are we affecting their values and attitudes? What kinds of citizens and what kinds of parents and spouses do they make? Are they becoming humane, more honest, and more concerned with the welfare of others? Are they becoming more active and better-informed participants in the democratic process? If we who are involved in the assessment game could succeed in persuading our faculty and administrators simply to begin seeking answers to such questions, we would be taking a major step toward institutional transcendence.

ASSESSMENT AND INCENTIVES: THE MEDIUM IS THE MESSAGE

An address by:

Linda Darling-Hammond
The Rand Corporation

Presented at:

The AAHE Assessment Forum
Third National Conference on Assessment
in Higher Education
June 8-11, 1988
Chicago, Illinois

Assessment and Incentives:

The Medium Is the Message

Linda Darling-Hammond

I come to the question of assessment as a teacher--as a former secondary school teacher, as a faculty member in an undergraduate institution, and a teacher in a graduate school of education. I see the issue of assessment as most important in its impact on the practice of teaching.

There are many other reasons why we conduct assessment, of course--reasons that this conference has attracted a thousand of you this year, a tremendous increase since a couple of years ago, when this issue was not very prominent on the higher education horizon. I suppose one of the reasons so many of you are here is suggested by the comment made by Mark Twain many years ago: "The legislature is in session, and no man is safe." The stimulus from statehouses and boardrooms to create assessment systems in colleges and universities is first and foremost a challenge to reframe the incentives that govern academic life. That is both its power and potential danger.

As you know, if there's one thing social science research has found consistently and unambiguously (and there are few things on which we social scientists agree), it's that people will do more of whatever they are evaluated on doing. What is measured will increase, and what is not measured will decrease. That's why assessment is such a powerful activity. It can not only measure, but change reality. One reason is that when assessment comes with stakes attached to it--stakes such as accreditation, student continuation in an institution, or faculty evaluation--it changes behavior and creates trade-offs.

HOW MEASUREMENT CHANGES BEHAVIOR

We can see this in other aspects of social life. Recently the FAA required airlines to begin to report how often they arrive on time at their destinations. The trade-off that poses, one acknowledged by that regulatory agency, is a trade-off with safety concerns. The question becomes how often, because they are evaluated on time-of-arrival, airlines will fail to make repairs that would make them run late. That potential problem has been partially taken into account in that particular assessment structure. But the point is that where you find an assessment measure, you will find a trade-off that has to be either taken into account or sacrificed to maximize a particular objective--in this case timeliness.

In American business, we have seen quite clearly the results of an emphasis on measuring short-term sales and profits, as against longer-range viability and market share. In fact, many analysts believe that American automobile manufacturers have lost market share to manufacturers in other countries because the measures that drove their operations, on which they gauged their performance, were not the long-run measures that would make them viable, productive industries over the long haul. Much of the change in business procedures in this country is an attempt to correct for what is now viewed as an overemphasis on inappropriate measures of success over the previous couple of decades.

In medicine, we have recently seen the publication of hospital mortality rates: the chances you have of dying if you go into a particular institution. Well, very quickly, we have begun to realize that that particular measure will produce incentives for hospitals not to care for very sick people. So, if we do not find ways to measure quality of care more appropriately, we will find institutions beginning to turn very sick people away because it will affect their mortality-rate measures.

The point is, in any arena, what you choose to measure and how you choose to use those measures will affect the very functioning of the institution. That is precisely the problem that you are engaging here today and will be grappling with in your institutions over the coming years.

I entitled this talk, "The Medium Is the Message" in remembrance of Marshall McLuhan's point that aside from the content of any communication, its form carries a meaning and effect all its own. So it is with educational assessment. What and how we look at what we do will, in itself, affect what we do by creating a set of incentives and disincentives for how we do it in the future. That can be powerfully positive. In some instances it can be powerfully negative.

For more than a decade, state policy makers have been reforming American elementary and secondary education, seeking,

especially, to make it more accountable. During a period when, for the most part (until the last couple of years), they were less willing or able to invest resources in the enterprise, assessment--especially student testing--has been a major part of that activity. And as the uses of these assessment data have grown, they have profoundly changed the shape and content of elementary and secondary education.

I think there are two implications of that activity for what is occurring in higher education today. One is that the forms of assessment that have been used for some years now in elementary education have, in part, produced the kinds of students who now enter your institutions. Students about whom state policy makers ring their hands. Students who, because they've been trained to take multiple-choice tests on a yearly basis, have stopped writing in their classes, have stopped reading books, have stopped engaging in complex problem solving or time-consuming laboratory tasks, and have come to college unprepared to engage in those activities. In fact, some of the events that have led to a concern for higher education accountability have been promulgated by the reforms of elementary and secondary education that have had a less than happy history. At the same time, the fact of reform at the K-12 level has produced a desire to reform higher education as well. If this new wave of reform takes the same course that the reforms of elementary and secondary education have taken, we may have lost the entire battle.

At the moment, the higher education enterprise in this country is well thought of the world over. Other countries pay to send their best and brightest to American colleges and universities to study and return home to increase the human intellectual capital in those countries. In many cases, those same countries consider the secondary and elementary systems in this country sub-standard compared to their own. What an irony. One of them, a gentleman from Britain, said to me, "Why don't you run elementary and secondary schools the way you run your colleges and universities, giving them a certain amount of control over how they conduct their work?" In fact, we're moving in the opposite direction: The policies that affect higher education increasingly resemble those that have already structured elementary and secondary education in this country. And, as I'll describe in a little while, some of those policies at which other countries turn up their noses have a great deal to do with assessment.

Now state policy makers have discovered higher education, and they're naturally extending their pet ideas to this sphere. Like assessment. In some cases they challenge you to assess yourselves. In others, they tell you how it should be done. In higher education, as in the K-12 arena, student testing is a key feature of many assessment plans. Though the same road is not yet taken, the issues are similar and the outcomes potentially the same.

So, what I'd like to discuss this afternoon is the difference between assessment as an improvement vehicle and

testing as an accountability tool. I'd like to place what is occurring in American higher education today in a context of assessment more generally, both here and in other countries.

INCENTIVES: A PARABLE WITH LESSONS

The importance of assessment can be demonstrated with a little story. The events I'm about to recount took place in Wonderland. And they happened something like this:

> Once upon a time in Wonderland, a prestigious national commission declared that the state of health care in that country was abominable. There were so many unhealthy people walking around that the commission declared the nation "at risk." Sweeping reforms were called for. In response, the major hospitals decided to institute performance measures of patient outcomes, tying decisions about patient treatment and dismissals to those measures.
>
> The most widely used instrument for assessing health in Wonderland was a simple tool that produced a single score with proven reliability. That instrument, called the thermometer, had the added advantage of being easy to administer and record. No one had to spend a great deal of time trying to decipher doctors' illegible handwriting or soliciting their subjective opinions about patient health. When doctors discovered that they would be held accountable for how many of their patients had abnormal thermometer scores, some complained that this was not a comprehensive measure of health. Their complaints were dismissed as defensive and self-serving. The administrators, to ensure that their efforts would not be subverted by recalcitrant doctors, then specified that subjective assessments of patient health would not be used in making decisions. Furthermore, any medicines or treatment tools that were not known directly to influence thermometer scores would no longer be purchased.
>
> After a year of operating under this new system, more patients were dismissed from the hospital with temperatures at or below normal. Prescriptions of aspirin had skyrocketed. And the use of other treatments had substantially declined. Many doctors had also left the hospital, arguing obtusely that their obligation to patients required them to pay more attention to other things than scores on the thermometer. Since the thermometer scores were the only measure that could be used to ascertain patient health, there was no way to argue if they were right or wrong.
>
> Some years later, during the centennial-year census in

> Wonderland, the census-takers discovered that the population had declined dramatically, and mortality rates had increased. As people in Wonderland were wont to do, they shook their heads and said, "Curiouser and curiouser." And they appointed another commission.

That little piece of history is being repeated in a number of different ways in American education. And it illustrates how it is that the choice of performance measures is extremely important when these measures become sanctions, incentives, and decision-making tools. In terms of assessment in higher education, there's a wide range of experience from across the states. There have been a number of positive effects already: a rebirth of accreditation and renewed attention to what it can mean, a renewed interest in comprehensive exams for students, surveys of graduates to find out what they're doing and what they have achieved since leaving college. All of those things can be very informative.

At the same time, some studies of state-mandated assessment programs, and particularly a recent one by Peter Ewell and Carol Boyer, have pointed out trends that make clear the importance of understanding the experiences of elementary and secondary education.

First, they found that state policy makers search for models. Because they have been through this with elementary and secondary education over the last 10 or 15 years, they often use the K-12 testing experience as a guide for what ought to be done in colleges and universities.

Second, assessment mandates, although they can be subject to a range of interpretations as to what forms of assessment are required, are often interpreted, either by the institutions or the regulators, as requiring common, standardized achievement testing--although a close look at the language of the mandate may not uphold that interpretation.

Third, what Boyer and Ewell call "the press to test" often absorbs the lion's share of assessment attention and resources, though tests may be one of only eight or ten different criteria or bases for assessment.

Fourth, assessment can prod introspection, evaluation, and improvement. Depending on how it's structured, it can also produce excessive paperwork, a narrowed curriculum, and a focus on ways to boost scores at the expense of real learning. Indeed, a recent study of the course of Tennessee's performance-funding initiative observed that, although a wide range of indicators were to be used and available, policy makers increasingly want to focus primarily on comparative test scores across institutions. This produces a range of incentives that can work against both educational improvement and equality of educational opportunity.

A lot depends, of course, on the forms of assessment, the uses made of assessment results, and the stakes that are associated with assessment. Over the last twelve or fifteen

years in elementary and secondary education, assessment has become, first, narrower in form and content with each successive state mandate; second, used for more purposes than those for which it was originally designed; and, third, tied in its results to greater and greater stakes with each successive refinement of state policy.

The same is likely to occur in higher education unless a new view of assessment is born and spread throughout this country.

THE K-12 EXPERIENCE: HIGH-STAKES ASSESSMENT

It's instructive to take a look at a brief history of what has happened in your sister institutions at the K-12 level. In the early 1970's, accountability legislation was passed in a great number of states, and it looked very much like the legislation that you're encountering today. It required planning and assessment systems; it was process-oriented. The institutions had to pledge to undergo some sort of internal evaluation and assessment.

Five or six years later, by the late 1970's, those methods had proved to be insufficient for policy makers' goals: Minimum competency tests were instituted. First, they were used as the criteria for graduating seniors. Later, the mandates were extended throughout the grades--with testing in grades 3, 5, 7, 9, and 12 in many states. A requirement that they be nationally-normed standardized tests became widespread. By then, the tests were to be used not only for graduation but for promotion, tracking, and sometimes (but not always) remediation of students. None of this provided what policy makers considered sufficient accountability.

And so, in the mid-1980's the states began to mandate curriculum guidelines that were heavily specified and aligned with the test. So, the current rage is curriculum alignment. The tests themselves didn't produce the outcomes desired, so now curricula are being redefined and mandated to match the test.

Part of the reason for this trend is the increased use of test results and their higher stakes. For example, standardized tests are now used as the basis for decisions about even such things as students' graduation from kindergarten in states like Georgia. They're used for decisions about remediation, tracking, placement in gifted and talented programs, and a variety of other things--things that these particular tests were not created or designed to do. In some states, they're now used as the basis for decisions about school funding. (We see that in the higher education realm in a few places.) Mandated tests are now used in elementary and secondary education as the basis for evaluating teachers for promotion, for career-ladder status, for salary, even for tenure and retention. They're used in some places for evaluating administrators.

We see in this brief account of the K-12 experience that

three things tend to happen when assessment measures are introduced. First, once a measure gains currency, policy makers have a tendency to use it for purposes and decisions not originally intended, as they arise, and regardless of whether the measure is intended for or suited to that purpose. Second, quantitative, comparative data, tend, over time, to override or overwhelm other forms of information, especially for people who are not expert in that area or enterprise. The tendency is to turn quickly and unskeptically to numbers, because, although we can't decipher all the other information about the quality of your physics programs, we all know that a 42 is better than a 38. Third, there's a tendency to forget what the number represents--if that was known to begin with. Robert Sternberg, a noted testing expert at Yale, makes the point that "the appearance of precision is no substitute for the fact of validity." Try telling that to a legislator who doesn't care to examine the disjuncture between the goals of your international studies program and the material on the ACT-COMP.

Once a measure is used to make high-stakes decisions about students, faculty, programs, or institutions, it can take on a life of its own. Walt Haney and George Madaus at the Center for Evaluation at Boston College make several points about high-stakes assessments, whether they are tests or other kinds of assessment.

First, the more any social indicator is used for social decision making, the more likely it will be to distort the social processes it is intended to monitor--in this case, an educational institution. Basically, any measurement of the status of an educational institution, no matter how well contrived, inevitably changes its status as people try to secure more of whatever is being measured.

Second, if important decisions are presumed to be related to test results, then teachers will teach to the test. This is easy enough to understand. High-stakes tests, it is argued, can, on the positive side, focus instruction, giving students and teachers specific goals to attain. Unfortunately, because such tests are indirect measures of the actual learning we care about, it's possible to do all kinds of things, quite successfully, to raise the test scores without actually increasing the amount of learning taking place. In fact, much recent research on the effects of high-stakes testing has shown that as scores on the instrument being used for assessment increase, scores on other measures tend to decline because of the shift in emphasis to that which is being measured. This has been studied and found to be so across tests, across settings, and in a number of countries. Studies of the effect of examinations on learning in Australia, India, Japan, Ireland, and England all turn up the same kind of result: that teaching to the test correlates with a de-emphasis on other forms of learning. So, we have to be very sure that what we're testing for is what we want, in fact, to promote, and that we don't, in fact, value other things as much or more than what we're measuring.

Third, teachers pay particular attention to the form of the questions on a high-stakes test--multiple choice, short answer, essay--and they adjust their instruction accordingly. The problem here is that the form of the test question can narrow instruction to the mode of the test as well as to its content. Haney and Madaus give this example from the Georgia Regents' testing program, which is designed to assess competencies of college students in reading and writing. The head of one English department in that state lamented:

> Because we are now devoting our best efforts to getting a larger number of students to pass the essay exam, we are teaching to the exam, with an entire course given over to developing one type of essay: a five-paragraph argumentative essay written under a time limit on a topic about which the author may or may not have any knowledge, ideas or personal opinions. Teaching this one useful writing skill has the effect of bringing a large number of weak students to a minimal level of literacy. But, at the same time, it devastates the larger purposes of the composition program, which should also be challenging students to produce a variety of different kinds of writing at more than a minimum level of competency. Because the Regents' Test is primarily designed to establish a minimal level of literacy, our teaching to the test, which its importance forces us to do, tends to make the minimal acceptable competency the goal of our institution--a circumstance that guarantees mediocrity.

In fact, however, Georgia is luckier than most. At least they are using an essay exam. In many places, and in K-12 education generally, multiple-choice tests mean that the incentive structure produces virtually no writing at all in classrooms. In fact, in most places now, particularly in large cities across this country, reading instruction has come to resemble the practice of taking a reading test. In reading class, students use commercial materials to decode short paragraphs about which they then answer multiple-choice questions. The teaching materials have evolved to resemble the tests the students will take. And tests dictate both the content of instruction and the teaching methods.

The fourth effect of high-stakes assessment is that, when test results are the sole or even the partial arbitrator of future life choices, society tends to treat them as the major goal of schooling rather than a useful indicator of achievement. So, ultimately, a student's score on the ACT-COMP, for example, may mean more to society than what she knows, more than anything else she has produced or done in the institution that she graduated from.

Finally, a high-stakes test transfers control over the

curriculum to the agency which sets or controls the examination. In this sense, obtaining control over the choice, content, and substance of whatever those measures are is the most important thing any institution can do to maintain control of its destiny. The phenomenon, as Haney and Madaus point out, is well understood in Europe, where systems of external certification exams controlled by central governments or independent examination boards operate at the secondary level. In this country, authority is increasingly being assumed for such decisions by state education agencies. And further, since states often use or mandate tests developed by outside test development companies, it is important to realize that the state may effectively be delegating this very real power to a commercial company whose interest is primarily financial and secondarily educational.

LIMITS OF STANDARDIZED TESTING

These five effects of high-stakes testing place a great responsibility on those who choose the measures. The coming of standardized achievement testing to higher education in America is a particular problem given the nature of testing in this country, which is very, very different from testing in most other countries. The United States invented McDonald's, and we invented multiple-choice, norm-referenced, standardized tests. We tend toward the quick, easy, and convenient, if less nourishing, approach to getting things done. But the kind of tests that are widely used in this country are rarely used abroad. And they are poor measures of most things that colleges are supposed to get students to do. In fact, if you think about it, the idea of a standardized, multiple-choice test of higher education is a stunning idea. What could such a thing be? Maybe a collegiate-level Trivial Pursuit, sampling those facts that a college educated person ought to know at the end of four years? Or maybe something that focuses on process... A set of general abilities that college is supposed to encourage? That means, in the parlance of psychometricians, that it probably measures the "G factor," that combination of general intelligence and test-taking skills which allows SAT, NTE, and GRE scores to correlate at a level of .9, irrespective of content and in virtually any sample in which you try to make that correlation. That doesn't tell you much about what a particular college or university produces in students, beyond what you could predict from knowing their SAT scores when they came in.

It's striking to think about such tests in contrast to other traditional forms of higher-education assessment: comprehensive examinations, theses, oral exams (which are still popular in some places, although less so than earlier), the use of outside examiners, demonstrations, exhibitions that are performance oriented...methods that in fact tell you whether somebody can do what they've supposedly been taught to do.

Indeed, many argue that it is the excessive use of narrowly defined, standardized testing in K-12 education that has produced some of the problems that colleges are now expected to solve:

students who come to you unable to write, think critically, solve problems, design and complete projects. The kinds of tests more frequently mandated require passive responses; they measure recall and recognition of facts, they aren't performance-oriented. And they can't represent either vast domains of knowledge well or higher-order thinking skills.

In contrast, in Europe and many parts of Asia, testing does not usually involve machine-gradable answer sheets filled out with number-two pencils. In England, students prepare written examinations in each of their areas of specialty. They also submit material from their portfolios of coursework. At the end of the secondary level in France, students take written and oral examinations in five areas. One of those, required of all students, is philosophy. Here are some sample questions from a recent exam: "What is judgment?" "Why should we defend the weak?" Compare that to what we might see in multiple-choice assessments in this country. In Germany teacher recommendations and grades, written examinations in German, foreign language, math, science, and social studies, and an oral exam are the basis for assessment. In Russia you might encounter an essay examination in Russian literature, and a series of oral examinations with two examiners in a variety of other areas.

And in many of these countries you would encounter, as well, a different notion of what assessment means to the structure of the educational profession. Faculties convene to develop the assessments. Furthermore, it is both the privilege and obligation of professors to examine the students of other professors at other institutions. The outside-examiner tradition in those places means that the act of assessment improves knowledge and cross-pollination across the enterprise as a whole, among professional faculty and the students. It is not the primary purpose of such assessment to produce and report two-digit data points to some other authority.

Recently, representatives from a group of countries came together to talk about international assessment and indicators. The United States made a proposal that the other countries adopt our National Assessment of Educational Progress (or NAEP) as the outcome measure. Not a single other country could be persuaded that doing so would in any way enhance the quality of their educational systems. In fact, the view most frequently heard was that assessment was too important to allow the adoption of a black-box measure that could drive instruction in directions we do not want to see it go.

It's time for educators in this country to begin to think about assessment as sufficiently important to heed such warnings. We must insist on creating and retaining control over measures of what we think a liberally educated person in this society needs to be able to know and do. We must not abdicate any of that responsibility to a measure that we don't thoroughly endorse.

There is a revolution in testing and assessment beginning to occur in this country. The kinds of concerns that I've been citing, primarily from the elementary and secondary domain, have begun to spur some action. ETS, for example, realizes that the days of the kind of test it has traditionally marketed are limited, and it's beginning to develop other forms of assessment. ETS president Greg Anrig has said, "the tests of the future will be individualized, open-response explorations of what students are learning over time." A new version of the National Teacher Examinations is already being devised for some of the same reasons. We need assessments which reflect performance. As I've mentioned, that's already the case in many other countries, where faculty are the ones developing the tests. They don't have to wait until a test development firm decides to do it. I think that is going to have to begin to happen here as well.

EFFECTS OF TESTING ON TEACHING AND LEARNING

We need better assessment because our tendency to treat tests as black boxes has become an increasingly serious problem. One example of how we're missing the mark is illustrated by the NAEP results, in which we have seen over the decade that students' abilities to understand basic concepts and principles, to analyze and make inferences, and to do problem solving have declined. The recent NAEP findings on reading achievement were accompanied by this commentary: "Only 5 to 10% of students can move beyond initial readings of a text. Most seem genuinely puzzled at requests to explain or defend their points of view." "Current methods," the NAEP assessors say, "of teaching and testing reading require short responses and lower-level cognitive thinking, resulting in an emphasis on shallow and superficial opinions at the expense of reasoned and disciplined thought." And given what most reading tests measure, it is not surprising that students haven't been taught more comprehensive thinking and analytic skills. And so it goes, in science, in mathematics, in history, in writing, and throughout the various subject areas.

Surveys have shown us that during the time state policy makers began to institute test-oriented accountability measures (between 1972 and 1980), the use of teaching methods appropriate to the teaching of higher-order skills declined in American public schools. There was a decline in methods such as student-centered discussions, writing essays or themes, and project or laboratory work. In 1980, fewer than 2/3 of high school students wrote regularly in any of their classes.

The National Science Foundation, The National Assessment of Educational Progress, and The National Councils of Teachers of English and Mathematics have all attributed this decline in students' problem-solving abilities to basic-skills testing in American schools. They charge that the emphasis on teaching what is tested in multiple-choice, standardized achievement tests has resulted in the neglect of higher-order thinking skills and performance abilities. There has been a de-emphasis on subjects that are not tested and on modes of performance other than

multiple choice, short answer, and fill-in-the-blank responses to predefined questions.

A number of recent studies uncover similar effects in colleges of education where standardized tests for graduation and licensure have been introduced, and where, in some cases, the pass rates of students on those tests have had implications for program approval. It is imperative that, as colleges and universities begin assessment activities, the difficulties that have undermined the health of the elementary and secondary system not be allowed to repeat themselves.

In fact, my hope is that more productive forms of assessment that are being developed already in many of your institutions filter down to the elementary and secondary schools. The caliber of those students who come to you from those schools will, after all, be in large measure a function of the caliber of assessment that goes on there. In one recent issue of Education Week, there were three headlines in the course of about as many pages that give an indication of what's happening to school curricula. One headline said "Using real books to teach reading said to heighten skill and interest." Why is this news? Well, as the article goes on to say, "the emphasis is on test-taking skills, not on learning reading or writing as a creative process." Another headline reads: "Best writing instruction uses all classroom resources and engages students in writing." In fact, a review of 72 studies shows that people learn how to write by writing! And not by being drilled on the rules of grammar. A third headline reports, "Study finds a neglect of humanities." The article indicates that half of the state officials surveyed attributed this neglect to the back-to-basics movement's reinforcement of teaching to standardized tests.

Ultimately, if we give the message that test scores are more important than learning, we'll see increases in test scores, but not necessarily in learning. Indeed, Lake Wobegon, where all the women are strong and all the children are above average, has already come to pass. A recent report released by the Friends of Education documented what some of us have been observing for some time: Every state in the nation now reports that its test scores are above the norm.

There are more important problems, though, when we think of using assessment, in the form of testing, as a public-policy tool. And I've seen this in some of the recently-enacted state policies to which some of you are subject.

POLICY MAKING AND ASSESSMENT

First, some policies define the institution as the unit of analysis, and then call for measuring "improvement" with some kind of average test score from year to year for that institution. But the institution is the wrong unit of analysis. It is meaningless to report that a school has, for instance,

improved its score. Schools do not take tests; students do. And students move into and out of institutions, carrying their test scores with them. You cannot infer institutional progress over time by looking at the average scores of students at one point in time and then the average scores of a different group of students at another point in time. To draw conclusions from such "snapshots" can lead to ridiculous conclusions about school "effects"--like that of the demographer who remarked about the amazing effect of living in Miami, where everyone is born Cuban and dies Jewish.

Furthermore, such invalid methods of comparison create very, very dangerous incentives with respect to equity. The best way to improve a school's average test score is to make sure the people who score poorly on the test don't come through the school. If funding is linked to test scores, schools will have incentives to push kids out rather than have them take tests on which they may do poorly. Incentives are created to reduce integration. To keep rich and poor students apart. To keep black and white students apart. To gain rewards, teachers and administrators will have to seek out schools like Lake Wobegon, where their value can be corroborated by the test scores of their students. The incentive is to affiliate with the students who need the least teaching, thus exacerbating what already tends to occur in the distribution of students (and resources to serve them) in this country. This situation is already embedded in policies that affect higher education in several states. But if we want to provide the public with real accountability, we have to attend to the distribution of resources and the quality of education--not the distribution of test scores.

To do this, we need to broaden the notion of what is a legitimate indicator of learning, looking not only at student test scores but at written and other types of performance as well; at student projects that demonstrate the ability to conceive an idea, work through problems, and produce solutions; at performances in journalism, the arts, debating, play writing; at research and demonstration projects in the natural and social sciences. Measures will have to become more complicated, involving real observations and judgments by teachers and administrators of what has been learned and accomplished. More external review, by faculties looking at each other's students, is called for--as is done in Europe. This is not to say that some paper-and-pencil test, even entailing standard questions, might not be appropriate. But let's think about formats other than multiple choice, and results that aren't automatically reduced to single numerical scores. Let's develop indicators of real performance of actual goals, rather than artificial measures of discrete sub-objectives that bear questionable relationship to the actual learning we value in our students.

The time is here for a new day in American assessment. And there are a lot of creative activities going on in a number of your institutions that can lead the way toward assessment that serves the goals of both accountability and quality education. A

new tradition of assessment requires us to be clear about what rewards and sanctions we're creating and how they relate to what we value and how we teach, to long-term learning and to equality of opportunity. Educationally sound assessment will mean paying attention to far more than the technical aspects of data collection and manipulation. It means allowing for relevance and coherence in each institution.

It also means paying attention to side effects, because whatever assessment schemes you adopt, your choices will have side effects. Watch for them, monitor them, and continually seek improvements. Educate the legislature if that is necessary. Educate those who are engaged with you in monitoring the assesment process. This is an extremely important activity. It can be extremely valuable. But if poorly done, it can be extremely harmful.

With the marriage of policy making and assessment, the work of educators and educational researchers has become more critical than ever before. We must insist on intellectual honesty and on educational validity. We must educate those who would impose hasty or inadequate methods. We must stand up for students, for academic quality and equality, and for a humane educational system. Professionals like yourselves who are rising to this challenge will make a profound and constructive contribution to American education.

THE ASSESSMENT MOVEMENT: WHAT NEXT? WHO CARES?

An address by:

Robert H. McCabe
Miami-Dade Community
College

Presented at:

The AAHE Assessment Forum
Third National Conference on Assessment
in Higher Education
June 8-11, 1988
Chicago, Illinois

Keynote Address
American Association for Higher Education
Third National Conference on Assessment in Higher Education
Chicago, Illinois - June 9, 1988

THE ASSESSMENT MOVEMENT: WHAT NEXT? WHO CARES?

by

Robert H. McCabe, President
Miami-Dade Community College

How has assessment reached the present high level of interest? What is going on in assessment? Where is the thrust coming from, and why?

Access and Standards

As background, it is relevant to review some of the events of American higher education in the 60s and 70s. The emergence of community colleges with an open door philosophy, fewer colleges and universities being really selective, and an increased interest by individuals in postsecondary education have resulted in a greater diversity of students.

There was substantial change in higher education in the 1960s. I left Miami in 1965 and went to New Jersey to open a new community college in downtown Newark. It was a great challenge and a good experience for me, and it calls to mind a key feature of the early 60s--it was a period of great optimism and stress on all systems. Americans were committed to progress in civil rights and believed that there would be rapid progress, that all of the shortcomings and the deprivation for minority groups were going to be overcome--and quickly. There was an

expectation of rapid advancement for minorities and other groups who had not been well served. In higher education, access was the message.

In community colleges we talked of "the right to fail." It was a "do-your-own-thing" period in which we believed that students knew more about what they should or could do than we did. And, therefore, we gave them permission to take any courses that they wanted. It might work out; if it didn't they had their "right to fail." In that process, and in the process of expanding access, we lowered expectations. I learned that very well when a faculty member in a campus outreach center told me, "I just can't ask these students under these circumstances, and with their personal commitments, for the same things I would ask students on campus." Across all of higher education, and particularly in open-door institutions, we did lower expectations. Despite the long term negative impact on higher education, in the context of those times, I would do the same thing I did in Newark in 1965 all over again. It was really important to get some of those individuals previously unserved through the system.

One major current problem is that many in education appear to have premitted their attitudes to become "stuck" in the 1960s. We haven't moved on as we should have from the emphasis on access to that of access with quality, the undeniable need for American society today.

At the same time these things were happening in higher education, something similar was happening in the schools, and the evidence of decline in the skills that students have as they graduate from our schools is very substantial.

That problem has been exacerbated by a rising literacy requirement. From the close of the Second World War to the beginning of the

1980s, the number of unskilled and semi-skilled jobs dropped from 80 to 20 percent. Thus, it is absolutely essential that larger percentages of Americans develop higher-level information skills so that they can participate effectively in society. I remember that when I graduated from high school with nine other individuals, I was the only one who went on to college, but there were jobs for all of the others. Today, that would not be the case. There are very few opportunities for individuals who do not have solid information skills.

With nearly two-thirds of high school graduates entering postsecondary programs within five years of high school graduation, and the growing number of high school graduates who are academically deficient, colleges design their programs to deal with a dramatically more diverse student body. They must also be prepared for a task new to many—helping underprepared students to academic success rather than weeding them out. In my opinion, the mismatch between the information skills of young Americans and those necessary to function effectively in society is the greatest threat to the well-being of the country, this side of a nuclear holocaust. Higher education has an important part in rectifying this situation by helping individuals who have completed high school, but do not have sufficient information skills, so that they are "salvaged." Assessment can contribute significantly to higher education efforts to address this problem.

The Public Call for Accountability

In recent years, the public has become increasingly critical of K-12 education and now higher education. There is a substantial amount of mistrust of teachers, and a growing dissatisfaction with the tremendous expenditure and seemingly limited results from our institutions.

The public wants assurance that something positive is happening. This is being reflected at the state level where legislators and state officials are telling educators, "We want you to perform in certain ways. To ensure that, we will regulate you more, and we want proof of performance." I remember in the 1960s when we would go to the legislature for appropriation. Our case was simple: "If you love your children, you should give us more money." And, legislatures responded with more money. Now there are discussions of relative merit and where limited resources should be spent. Ninety billion dollars goes into higher education annually in this country, and the public wants to know whether it is being well spent. And, as independent colleges and universities begin to draw more public funds, state and federal, they will not be exempt.

The response to recognition of unsatisfactory educational quality is coming from the public through legislatures and from institutions--both involve assessment. However, they think of assessment rather differently. At the state level, the assessment movement started as a standardized test movement with the K-12 system. Greg Anrig, the president of the Educational Testing Service, cites an old army saying: "If it moves salute it," and a current trend among state agencies is: "If it moves, test it." In fact, there are 24 states in which there is a test as a condition for graduation from high school. It is not something that is going to go away. There is more and more use of assessment through standardized tests occurring throughout the educational system--some controlling individual student progress, some focusing on the performance of the institutions. In the 1980s, this trend that began in K-12 is moving into higher education, and particularly in open-door

institutions. Community colleges are certainly the most vulnerable to poorly conceived programs because these institutions have the most diverse student bodies. In the push for quality, many question the ability of these institutions to achieve a quality outcome with so many underprepared among their students. Others question the efficacy of expenditures for second and third chances for students whom they believe had their chance in high school. What those of us in the open door colleges learn is that the fact that an individual does not have sufficient academic competence does not mean there is necessarily lack of talent, or that the talent cannot be developed.

State Initiatives in Assessment

Two-thirds of states now require some type of assessment in public colleges and universities. There are two primary issues at the state level. One is whether the assessment is going to be of the student, that is, as a condition for the student to move from one level to the next, or whether its purpose is to evaluate or to give feedback for program improvement. The second is whether the state designs and specifies the assessment program, or whether institutions are permitted to devise their own programs.

It appears that there is a strong trend among state authorities to require assessment, but to consider the design and conduct of assessment a matter of institutional prerogative. There are very good reasons why that should be the case. However, as I meet legislators, governors, and state officials, they express a real sense of impatience--a feeling that institutions will not do anything unless the state forces them to do so. Therefore, while most state officials are saying, "Assessment is an activity in which institutions should take the lead, there is a deep

suspicion that the institutions aren't going to do it--we're going to have to do it ourselves." And, when it comes to educational program, legislators and state officials are much more willing to do it themselves, as they have through a growing array of rules affecting education, including specifications on how certain courses are to be taught, and what the content of the courses will be. We are seeing much more regulation and movement to assessment based on a growing mistrust of college educators.

Not surprisingly, the states rely heavily on standardized tests. My feeling is that the less one knows about standardized tests, the more willing one is to rely on them. (Which is not to say that I don't see their value.) The more one knows, the more one is aware of the limitations. The less one knows, the more appeal of "scientific" measurement that can be objectively expressed. State officials turn to standardized tests because they are objective, with numbers that can be used for comparisons or standards. They provide an easy way to look at reading, writing, and mathematics skills, which is where much of the state thrust is aimed.

What concerns me most is the use of standardized tests by states as a single measure of performance. It is one thing for the states to say, "We are giving you tremendous amounts of money--we want information to indicate whether you are being successful." It is quite another for the state to mandate a particular test or tests to determine when they will be given, who will administer them, and what will happen with the results.

The Florida Experience

Let me turn to my own state, Florida, as an example of the extreme side of the state-based, standardized test-oriented approach to assessment. I am going to be very critical, but I want to say up front that the project has had some positive impact. There is no doubt that the Florida higher education institutions are more concerned about student information skills, and there is more effort in placement and remediation. However, on the whole, it is my opinion that the Florida approach is not the right one. It should be modified in Florida and not copied in other states.

Let's begin with the first part of the Florida program, a statewide, standardized test for placement. Legislators are comfortable with this requirement because it allows them to say: "Okay, now we have uniform data. We know whether students are really academically deficient or not." Without doubt, mistrust is part of the appeal of the statewide placement testing as it provides a standard which won't allow institutions to "cheat." Cheat, that is, with regard how they are funded and how they report enrollment. This approach has two major disadvantages. First of all, each of our institutions has a different curriculum--and we certainly want to keep it that way. The use of a placement test should be tied to the curriculum and its special features. For example, at my institution, our approach to placement (which we were doing before the state mandated it) was to determine placement only after we examined the relationship between performance on the test and performance in college general education courses. That is, we based placement on what we knew about student performance in our curriculum. Second, once students were placed, the faculty could adjust placement

based on classroom performance. For example, in writing courses, on the first day of class students wrote an essay, and on that basis faculty could move them to a more appropriate class.

The point is that placement testing is more likely to benefit students when it is designed and administered by the institution. It should be tied to the curriculum; its primary purpose should not be reporting to the state, but improving the growth and development of students. The college has the advantage of using faculty judgment and classroom performance for classroom placement, and can follow up on the programs of placement and remediation to determine if they have the intended effect.

The next piece in the Florida program is the rising-junior test: the College Level Academic Skills Test (CLAST), required of all students in public institutions (or in independent institutions receiving state aid) for entrance to upper level coursework or the award of an associate in arts degree. This is the most controversial component of the testing program. The CLAST experience shows the concerns that all of us should have about standardized rising-junior examinations. One concern is that it drives curriculum. In the case of this examination, which was designed by faculty from across the state, there are five subdisciplines in the mathematics test. Therefore, the colleges have had to redesign the curriculum in order to align with that particular set of mathematics competencies. It appears that many of those competencies are not needed by many of the students, but their curriculums by necessity include a sequence of courses to learn these competencies. A similar problem occurs in the area of composition. In my institution, half of our 45,000 credit students have a native language other than English. You

can imagine the task of getting these students through the composition section of CLAST--particularly as it is timed, and it may be on a topic with which they have no familiarity. It makes it very tempting (and I heard this proposed on one of our campuses) to turn the English curriculum into a program to teach students how to produce a credible essay within extreme time constraints--simply test preparation. And this flies in the face of what most faculty tell me they should be teaching. They tell me that they teach students to organize ideas, to outline, to revise, and to use dictionaries. The need to help students pass CLAST drives us toward a curriculum that the faculty do not support. This is certainly not in the best interests of students. Further, the increase of English and mathematics enrollment is squeezing students out of sophomore level courses in the humanities, sciences, and social sciences. CLAST is beginning to dominate curriculum.

Most important is the unquestioned use of CLAST as an independent criterion of student success. Is a statewide standardized test a valid predictor of students' ability to succeed? The fact is, there are large numbers of students who did not pass CLAST and proceeded to the upper division (when that was permitted), and are performing well at the junior and senior year. Studies suggest that the best forecaster of success is a combination of grades and a standardized test, the next best is grades, and the least effective is a standardized test by itself.

CLAST is having a particularly devastating impact on minorities. The number proceeding to upper division is in sharp decline, and our data suggest that many of those could be successful. So why not consider grades along with test scores? One must ask, in a country where there is a severe problem with the small number of minorities advancing

through each level of education, why would a program be utilized that is cutting out many of those who could succeed from proceeding through baccalaureate programs. It simply doesn't make any sense.

Unfortunately, public relations plays an important part in statewide test programs. It has impact on the public image of the institutions, and can result in decisions based on that rather than what is good for students. I know of two institutions in Florida that are screening students before they take the CLAST. So, in fact, only those students whom they are convinced will pass take the test. The result when the newspaper article comes out and ranks institutions by results, those two institutions look good. But is that good for students? How about the student who has a good, though not sure, shot at passing? That student ought to have the opportunity to take the test. The question shouldn't be "what's going to give us the best public image," but "what is in the best interest of the student?" A year ago the legislature passed a rule which permitted students to take the CLAST at any time. That made sense as students who had the competencies at admission or early in their student careers, could be exempted for the CLAST curriculum sequence and permitted to enroll in a richer curriculum. One of my colleagues argued against this. "Wait a minute," he said, "if we do that, the pass percentages on CLAST for my institution aren't going to look as good." And frankly, when institutions are ranked in the newspaper four times a year on the basis of scores, it is hard not to think that way. It is a serious problem.

And so I hope we can make some changes in the Florida program, not do away with it, but make changes that will make it more beneficial to students. We need to keep in view two things that distinguish American

higher education from other systems. One is institutional and curricular diversity, which is threatened by the imposition of a standardized test; and the other is giving second and third chances to students, which is threatened by making judgments about students on the basis of a single criterion--a standardized test.

Finally, we come to an issue that underlies all the others--the continuing thrust toward centralization. Everything I read about management indicates the importance of making decisions as close to the action as possible, and involving people in decisions that impact them. What Florida is doing in assessment runs contrary to good management practice. The program does not tie to the curriculum. It is not designed to give feedback to students, and it bypasses faculty. Basing decisions on a single statewide measure sets aside all of the expert judgment that faculty have exercised over the years in the classroom. And I think in some cases it can force students into poor practices. Florida has taken leadership in utilization of statewide assessment, and much has been learned--positive and negative. Now we need to utilize our experience and take responsibility to design an assessment program that helps students learn and grow. At the same time, we must recognize that the public has every reason to expect us to do more and to show the results of our efforts.

Institutional Assessment Initiatives

I would like to turn now to the other aspect of what is happening in assessment--institutional initiatives. They have great potential. Much of the impetus here comes from institutions looking at themselves, their programs, and results, and asking, "Do we really know what we're

trying to achieve?", and "Have we got any way of determining whether we are being successful?"

There may have been a time when one of the roles of universities and colleges was to screen out students who were not prepared for immediate success. Full responsibility was on the student. We all know the stories of the dean addressing freshmen at orientation and saying, "Those of you who don't find the library by the end of the first football game won't be here by Thanksgiving." Underlying this attitude was an assumption that entering students had been well-prepared and should be able to take care of themselves. I think that approach is history, for all but a few institutions. When we look at the students who are entering colleges today, and consider the needs of the society for increased numbers of people who have strong information (academic) skills, it is clear that institutions must take greater responsibility for student success. The job is no longer to screen out, but to help more students, including the underprepared, to quality academic performance. Placement and program guidance are now important concerns, so there is good reason to develop assessment programs. Not only does the public have a right to know how well we are doing, it is vital to students that we know more about them. The idea that students either get it or get out just isn't appropriate in the face of massive underpreparation and growing diversity. This nation needs to develop all of its talent, and the fact that an individual is not well-prepared academically at one point in life does not mean that he or she has no talent or potential. Assessment can help us tap that talent and potential.

For assessment to help it must involve faculty. When we assess the effectiveness of the curriculum, the faculty must participate in deter-

mining those desired program outcomes. There must be overall purpose and unity to the curriculum. The major hurdle is for faculty to understand that the courses they teach are not entirely independent. That is going to be very difficult to achieve, but it is essential if assessment is going to make any difference in improving student programs. The key is to understand that the focus is on the student and what the courses can contribute to his or her development. Moreover, I would argue that we have some obligation to assess teaching itself. Colleges have the reverse of the problem in K-12, where the concern is that the faculty do not have adequate mastery of their subject matter. In higher education faculty are well grounded in their disciplines, but they don't really know a great deal about teaching or learning. This is particularly striking when considered in light of the significant body of knowledge about teaching and learning that has accumulated in the past thirty years--knowledge that is not a part of the vocabulary of our faculty. Most come to college teaching with a love for their subject matter, and a love for the kind of work they did in graduate school. But teaching is something they learn on the job or not at all. A medical analogy is suggested. If I needed heart surgery, I would look for a doctor who knew more about the heart than anybody else in the world. But if that person had never had a scalpel in hand, I'm not sure I'd want to be operated on. And to a great extent, this is the situation with faculty. They have wonderful knowledge about their special field, but little knowledge of "the operation"--teaching and learning. College teaching is one of the few professions where we don't stand on the shoulders of those who went before, learning from them, so each generation improves. That really needs to change, especially as institutions and faculty take

greater responsibility for student performance. It is time to come to grips with what is expected in faculty performance, and to think about that in light of what assessment can tell us about student learning. We need to help faculty become better teachers, and we need to help them assess how well they succeed.

Related to this is shifting concern to our output instead of input. What do colleges brag about? The SAT scores of incoming students--how many students came with this or that level of ability. Most have little to say about how students grew or what they can do as a result of our programs. I was at a meeting of state leaders in Florida. They were discussing a suggestion that our goal should be to improve the quality of higher education in Florida. I stated my belief that this was an inappropriate goal. Why? Because it would be very easy to improve the quality of higher education by simply reducing the number of students admitted by screening out the less well-prepared. Presto! The equality of higher education would be enhanced. But the goal must be for education to be more successful in meeting the needs of society. If you put it that way, you come up with a very different kind of solution. Improving quality involves expanding the number of students gaining academic achievement. To realize this goal, colleges should assess the effects of educational programs on students and develop methods to get that information back to students in ways that help them know how they are doing, what they can do better, and what their next step needs to be. The same information needs to be available to each teaching faculty member as a basis for improving teaching.

This, it seems to me, is what Pat Cross has been talking about--classroom research. I am very impressed with the potential at the level

of assessment she advocates--assessment that gets down to what each faculty member does with students--the place where the real learning takes place.

Summary

Where are we and what has happened with assessment? I have commented about state initiatives and about institutional initiatives, and some of the issues that arise in the tension between those initiatives. Should assessment be the responsibility of the state or each institution? Is it to make determinations about individual students or to evaluate the institution? Is it for both? Should it be standard across institutions, or diverse? How should the results be utilized?

Where do we go from here? There is little doubt that this assessment movement is going to grow. There is no doubt that the interest of legislators and the public is increasing. And I am encouraged by the growth of interest that I see within institutions, particularly geared to student learning and student growth. Hopefully that is where the most energetic assessment efforts will occur. Assessment should be an ongoing vehicle for self-awareness and change.

The assessment movement is growing in tandem with the teaching/learning movement. In fact, it could be considered an element of it. There is tremendous interest in improving teaching and doing more to improve to improve student learning. Assessment is essential to that, whether it be classroom research or program assessment. The most promising assessment programs now in progress deal with the impact of teaching on student learning and feedback to students and faculty. Lee Shulman's work comes to mind, as does the program at Alverno College. What I am suggesting is that the future of assessment will be in improv-

ing student development through more effective teaching and learning. In the 1960s community colleges operated on the basis that students had a "right to fail." We should not operate on the basis of their having a "right to succeed," and assessment can contribute to that success.

ADDITIONAL RESOURCES
AAHE ASSESSMENT FORUM

The following resources are available for purchase from the AAHE Assessment Forum:

1. Resource Packet I: Five Papers **$15.00**

--"Assessment, Accountability, and Improvement: Managing the Contradiction," P. Ewell
--"Assessment and Outcomes Measurement: A View from the States," C. Boyer, P. Ewell, J. Finney, and J. Mingle
--"The External Examiner Approach to Assessment," B. Fong
--"Six Stories: Implementing Successful Assessment," P. Hutchings
--"Thinking About Assessment: Perspectives for Presidents and Chief Academic Officers," E. El-Khawas and J. Rossmann

2. Resource Packet II: Six Papers **$25.00**

--"Acting Out State-Mandated Assessment: Evidence from Five States," C. Boyer and P. Ewell
--"Assessing Student Learning in Light of How Students Learn," J. Novak and D. Ridley
--"Faculty Voices on Assessment: Expanding the Conversation," P. Hutchings and E. Reuben
--"Feedback in the Classroom: Making Assessment Matter," K. Patricia Cross
--"Standardized Tests and the Purposes of Assessment," J. Heffernan
--"An Update on Assessment," (AAHE Bulletin, December, 1987), P. Hutchings and T. Marchese

3. Three Presentations--1987: **$8.00**

from the 2nd National Conference on Assessment in Higher Education
--Lee S. Shulman -- "Assessing Content and Process: Challenges for the New Assessments"
--Virginia B. Smith -- "In the Eye of the Beholder: Perspectives on Quality"
--Donald M. Stewart -- "The Ethics of Assessment"

4. Three Presentations--1988: **$10.00**

from the 3rd National Conference on Assessment in Higher Education
--Alexander Astin -- "Assessment and Human Values: Confessions of a Reformed Number Cruncher"
--Linda Darling-Hammond -- "Assessment and Incentives: The Medium is the Message"
--Robert H. McCabe -- "The Assessment Movement: What Next? Who Cares?"

5. Assessment Programs and Projects: A Directory **$10.00**

Concise descriptions of thirty assessment projects being implemented on campuses across the country. Edited by Jacqueline Paskow.

To order items indicated above, contact: Patricia Hutchings, Director, AAHE Assessment Forum, One Dupont Circle, Suite 600, Washington, DC 20036; 202/293-6440. Orders under $25 must be prepaid. Allow four weeks for delivery. MAKE CHECKS PAYABLE TO AAHE ASSESSMENT FORUM.